

Mayfield High School - Intermediate Tier

Mayfield High is a fictitious school but the data collected is from real students. You will investigate the relationship between height and weight. There are mistakes in the data including impossible heights and weights.

Locating the Mayfield Database

In the school Intranet select **Subject and Group Sites**

Then select **Maths Website**, followed by **Mayfield** and click on the word **HERE**

You will spend some time on the computers obtaining data to work on and producing graphs and statistical calculations, but marks are awarded for your **entire write up** of the coursework. Your report must contain the following:

1. A detailed plan for your study
2. Samples obtained
3. Calculations and graphs produced
4. Detailed conclusions
5. An evaluation

Firstly, as you are only interested in height, weight, year and gender you must delete the rest of the data.

To Delete Unwanted Data (leaving Year, Gender, Height and Weight):

Select columns 2 to 7

From the **Edit** menu select **Delete**

Now repeat this for columns 3 to 12 and 5 to 11

REMEMBER TO REGULARLY SAVE YOUR WORK INTO YOUR **MY WORK** FOLDER

You will be testing the following hypotheses:

1. **The taller you are the heavier you are**
2. **Boys are taller and heavier than girls**

You need to say how you think age and gender will affect results.

e.g.

Hypothesis 1 - The taller you are the heavier you are. I think the relationship between height and weight will become stronger as you get older.

Hypothesis 2 - Boys are taller and heavier than girls. I think the difference between boys and girls will increase as the students get older.

Taking a Sample

There are the 1183 students at Mayfield High School. The table shows the number of students in each year group for males and females.

Year Group	Boys	Girls	Total
Year 7	151	131	282
Year 8	145	125	270
Year 9	118	143	261
Year 10	106	94	200
Year 11	84	86	170

For your study you will deal with each of the 10 groups in the school (i.e Year 7 Girls, Year 7 Boys, Year 8 Girls etc) separately to make comparisons across year groups and gender. To do this you will need samples from each group as the whole school is too large to work with. 30 students from each group should be large enough to perform your calculations on.

Random Sample - Every person should have an equal chance of being chosen for your sample to make it fair and avoid bias. A quick way of doing this is to give each student a random number and then sort the data on this number which produces a random list of the data.

To Get a Random Sample in Excel:

1. Type **=RAND()** in the first free cell to the right of the first line of data and press **Enter** to insert a random number.
2. Click on this cell again and move the cursor to the bottom right of the cell until it changes to a black cross. Drag down until you reach the bottom of the data.
3. To mix up the data, highlight the cell to the right of the first random number. Select the **Data** menu and **Sort**. **Sort by column 5** and this will mix up all the data.
4. Now select **Data** menu and **Sort** and **Sort by Year Group Then by Gender**.
5. Select the number of calculated students from each group and copy to a separate sheet

Calculations And Evidence Required For Your Study

1. **Scatter graph** for each group
2. **Correlation coefficient** for each group
3. **Line of best fit** and **equation** on scatter graphs (only if correlation coefficient is >0.6)
4. **Minimum, lower quartile, median, upper quartile** and **maximum** for weights (or heights). From this you can find the range and interquartile range and produce **box and whisker diagrams** (these should be done by hand on graph paper)
5. **Mean** and **range** for heights (or weights) for each group

Anomalies

After drawing your scatter graph you may notice anomalies (extreme values in your data, like someone who is 4.65m tall or 5kg in weight). These anomalies may be outliers (i.e. results differing greatly from others in the same sample). You should decide how to deal with them (i.e. leave them in or replace them).

Replacing Anomalies

If you find an outlier in your data you must deal with it and include evidence. If you decide it is an impossible piece of data and remove it, then you should replace it with a new person from your original randomised data. Your scatter graph and calculations will automatically update, so remember to print out any graphs and calculations before deleting to use as evidence in your report. Explain your reasons for taking them out and explain how anomalies could affect your results.

Producing Calculations and Graphs

Scatter Graphs

These are used to compare the relationship (correlation) between two types of data.

Correlation coefficient

This is a more accurate method to compare correlation. It uses the mean of each set of data and looks at the distance away from the mean of each point.

The value determines correlation. It is always between -1 and 1.

-1 = Perfect Negative Correlation	1 = Perfect Positive Correlation
-0.8= Good Negative Correlation	0.8= Good Positive Correlation
-0.5= Some Negative Correlation	0.5= Some Positive Correlation
0 = No Correlation	

A line of best fit should only be drawn on a scatter graph if the correlation coefficient is >0.6 or <-0.6

The reason for displaying the equation of line of best fit is that it can be used to make predictions.

E.g. If the equation for Year 8 boys is $y = 50x - 40$ this means for a boy in Year 8 his weight can be predicted if you know his height from calculating
 $\text{weight} = 50 \times \text{height} - 40$.

To Draw a Scatter Graph in Excel:

1. Highlight the two columns of data
2. Click on Chart Wizard (Bar chart icon on tool bar)
3. Choose **XY(Scatter)**
4. Enter chart title and label axes(remember units!)
5. In **Legend** untick box labelled **Show Legend**
6. Choose whether to save as separate chart or on sheet

To Improve Presentation:

Right click on x-axis and select **format axis**, choose **scale** and change **minimum value**. Can repeat for y-axis if necessary.

To Put on a Line of Best Fit (only if strong enough correlation):

Right click on a point in the scatter graph, select **add trendline**. In **options** tick box to **display equation**.

To Calculate Correlation Coefficient in Excel:

1. Select a blank cell in the spreadsheet
2. Click on **fx** on the tool bar
3. Select **Statistical** in the **function category**
4. Select **Correl** in the **function name** and then click **ok**
5. In **Array 1** highlight the heights
6. Click in **Array 2** and highlight all the weights
7. Click **ok**

Box and whisker diagrams

This is a visual representation of the minimum, lower quartile, median, upper quartile and maximum. From box plots you can compare the medians (middle value), range and interquartile range (middle 50% of data found by subtracting the lower quartile from the upper quartile)

To Calculate Quartiles in Excel:

The Lower Quartile

1. Click in an empty cell
2. Click on **fx** on the tool bar
3. Select **Statistical** in the **function category**
4. Select **Quartile** in the **function name** and then click **ok**
5. Highlight the column of data, which will appear in the **array** box
6. Click in **quart** box and type 1
7. Click **ok**

To calculate the Minimum, repeat as above but type in 0 instead of 1 in the **quart** box

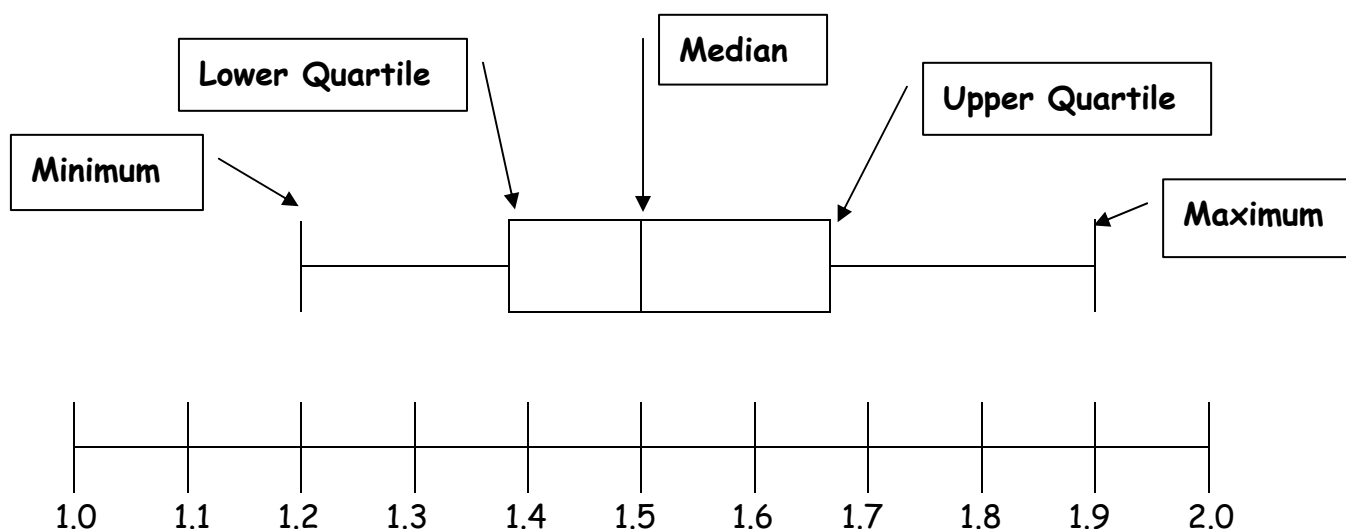
To calculate the Median, repeat as above but type in 2 in the **quart** box

To calculate the Upper Quartile, repeat as above but type in 3 in the **quart** box

To calculate the Maximum, repeat as above but type in 4 in the **quart** box

Drawing Box and Whisker Diagrams

1. Use graph paper
2. Draw a horizontal scale. It is a good idea to draw all the box plots on one piece of graph paper using the same scale. You will then be able to compare your results in your conclusion.
3. Mark on the 5 pieces of data you have found and draw in the box and whiskers as below.



Mean and Range

The mean is a type of average, the range is a measure of spread.

$$\text{mean} = \text{total of values} \div \text{no. of values}$$

$$\text{range} = \text{biggest value} - \text{smallest value}$$

To Calculate the Mean in Excel:

1. Click in a blank cell
2. Click on **fx** on the tool bar
3. Select **Statistical** in the function category
4. Select **Average** in the function name and then click **ok**
5. Highlight the list of numbers you require the mean for, which will appear in the **number 1** box (ignore **number 2** box) and click **ok**

To calculate the range, repeat as with the mean, but click on **min** in the function name column instead of average. Then repeat but click on **max** in the function name column. To find the range subtract the min value from the max value.

Summary Tables

Summary tables are used to make comparisons between years and gender easier. You should include key pieces of data in your summary table(s):

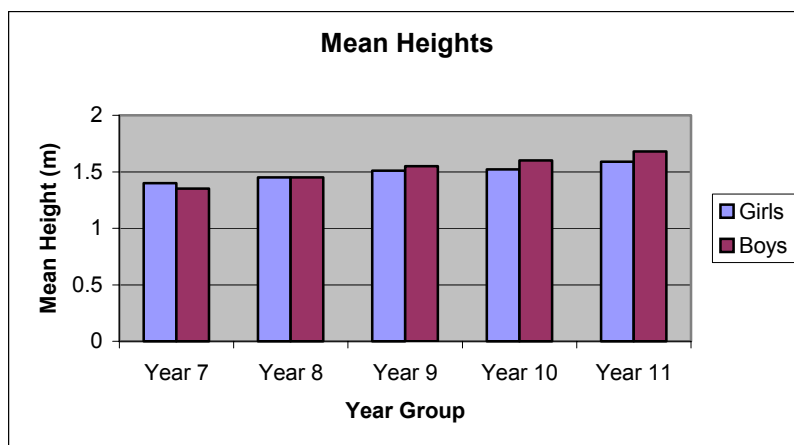
- **Correlation Coefficient**
- **Mean and Range**
- **Median and Interquartile Range (IQR)**

A possible summary table layout:

Group		Correlation Coefficient	Height		Weight	
			Mean	Range	Median	IQR
Year 7	Male					
	Female					
Year 8	Male					
	Female					
Year 9	Male					
	Female					
Year 10	Male					
	Female					
Year 11	Male					
	Female					

Bar charts can also be a way to compare these values visually.

E.g.



Conclusions

Your conclusions **very** important and must be detailed, but clear and not waffle! Any calculations which you do not mention are considered irrelevant data. Make statements and comparisons from your calculations across year groups and gender and link these to your original hypotheses. (More info in *Your Report* section of this booklet)

Evaluation

You must consider how reliable your findings are. Can you use them to make correct predictions for the whole school? Could they be used to make predictions for your school?

One way of considering the limitations of your sample is to compare the key results in your summary table to calculations for the whole population (use Excel to obtain these calculations)

Your Report

Your write up must include the following:

Plan

- ❑ Brief introduction to the coursework task
- ❑ State hypotheses (to include variation across years and gender)
- ❑ How you will prove/disprove your hypotheses (i.e. What calculations and analysis you will make- Explain which calculations will prove which hypothesis and WHY you use them)
- ❑ Why/how sample is taken; How many; What groupings
- ❑ How you will eliminate bias/deal with outliers (include evidence)

Represent

- ❑ Scatter graphs for each group, correlation coefficient, line of best fit and equation (where appropriate)
- ❑ Calculate min, max, lower and upper quartiles and median for heights or weights of each group
- ❑ Box and Whisker diagrams (hand drawn on same sheet for later analysis)
- ❑ Summary table of calculated values

Conclusion

- ❑ Correlation shown on each scatter graph and what they evidence
- ❑ Comparison of correlation values across the groups
- ❑ Explain equations of lines of best fit and use to make predictions/inferences
- ❑ Comparison of height/weight between boys/girls
- ❑ Comparison of height/weight across year groups
- ❑ Refer to box and whisker diagrams and what they demonstrate
- ❑ Relate findings to original hypotheses. Do your findings support them?
- ❑ Give/suggest reasons for your findings
- ❑ Give/suggest reasons for exceptions to patterns in your data

Evaluation

- ❑ Was your sample a good/fair/representative sample? (Compare to whole school results?)
- ❑ Any limitations of the project?
- ❑ Any problems you faced and changes you would make